Math 164: Problem Set 8

Owen Jones

3/1/2024

10.10 (a) $f(x) = \frac{1}{2}\mathbf{x}^\top \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \mathbf{x} - \mathbf{x}^\top \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

(b) $\mathbf{d}^{(0)} = -\mathbf{g}^{(0)} = -\begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \mathbf{x}^{(0)} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$,

$\alpha_0 = \frac{5}{29}$,

$\mathbf{x}^{(1)} = \begin{bmatrix} \frac{15}{29} \\ \frac{5}{29} \end{bmatrix}$

$\mathbf{g}^{(1)} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \frac{15}{29} \\ \frac{5}{29} \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{29} \\ \frac{6}{29} \end{bmatrix}$,

$\beta_0 = \frac{\mathbf{g}^{(1)\top}\mathbf{Q}\mathbf{d}^{(0)}}{\mathbf{d}^{(0)\top}\mathbf{Q}\mathbf{d}^{(0)}} = \frac{4}{841}$

$\mathbf{d}^{(1)} = -\mathbf{g}^{(1)} + \beta_k \mathbf{d}^{(0)} = \begin{bmatrix} \frac{2}{29} \\ -\frac{6}{29} \end{bmatrix} + \frac{4}{841} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{70}{841} \\ -\frac{170}{841} \end{bmatrix}$

$\alpha_1 = \frac{\frac{1160}{24389}}{\frac{5800}{841^2}} = \frac{29}{5}$

$\mathbf{x}^{(2)} = \begin{bmatrix} \frac{15}{29} \\ \frac{5}{29} \end{bmatrix} + \frac{29}{5} \begin{bmatrix} \frac{70}{841} \\ -\frac{170}{841} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(c) By the FONC $\mathbf{x}^* = \mathbf{Q}^{-1}\mathbf{b} = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ which agrees with the result from part (b).

12.1 $\arg\min_m \|\mathbf{A}m - \mathbf{F}\|^2$ where $\mathbf{A} = \begin{bmatrix} 3 \\ 5 \\ 6 \end{bmatrix}$, $\mathbf{F} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. $m^* = \mathbf{A}^\top\mathbf{A}^{-1}\mathbf{A}^\top\mathbf{F} = \frac{31}{70}$

12.8 $y_0 = 0$
$y_1 = bu_1 + v_1$
$y_2 = abu_1 + av_1 + bu_2 + v_2$
$\cdots$
$$y_k = \sum_{i=1}^{k} a^{k-i}bu_i + a^{k-i}v_i$$

Let $\mathbf{C} := \begin{bmatrix} b & 0 & \cdots & 0 \\ ba & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ ba^{n-1} & \cdots & ab & b \end{bmatrix}$ and $\mathbf{D} := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a^{n-1} & \cdots & a & 1 \end{bmatrix}$

$\mathbf{y} = \mathbf{Cu} + \mathbf{Dv}$

Matrices $\mathbf{C}$ and $\mathbf{D}$ are diagonal matrices, so their determinants are the products of their diagonals. Their determinants are non-zero, so they are nonsingular. Let $\mathbf{b} = \mathbf{D}^{-1}\mathbf{y}$ and $\mathbf{A} = \mathbf{D}^{-1}\mathbf{C}$. $\mathbf{C} = b\mathbf{D}$, so $\mathbf{A} = b\mathbf{I}_n$. Thus, the linear least square estimate of $\mathbf{u}^* = \frac{1}{b}\mathbf{D}^{-1}\mathbf{y} = \frac{1}{b}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -a & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -a & 1 \end{bmatrix}\mathbf{y}$

(using row reduction to solve for $\mathbf{D}^{-1}$)

12.10 (a) Using $\arcsin(y_i) = \arcsin(\sin(\omega t_i + \theta)) = \omega t_i + \theta$ for $-\frac{\pi}{2} \leq \omega t_i + \theta \leq \frac{\pi}{2}$, we obtain the following system of equations:

$$\begin{aligned} \arcsin(y_1) &= \omega t_1 + \theta \\ \arcsin(y_2) &= \omega t_2 + \theta \\ &\vdots \\ \arcsin(y_p) &= \omega t_p + \theta \end{aligned}$$

(b) Let $\mathbf{A} := \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_p & 1 \end{bmatrix}$, $\mathbf{x} := \begin{bmatrix} \omega \\ \theta \end{bmatrix}$, and $\mathbf{b} := \begin{bmatrix} \arcsin(y_1) \\ \arcsin(y_2) \\ \vdots \\ \arcsin(y_p) \end{bmatrix}$

It follows $\mathbf{x}^* = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{b} = \begin{bmatrix} p\overline{T^2} & p\overline{T} \\ p\overline{T} & p \end{bmatrix}^{-1} \begin{bmatrix} p\overline{TY} \\ p\overline{Y} \end{bmatrix}$

$= \begin{bmatrix} \overline{T^2} & \overline{T} \\ \overline{T} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \overline{TY} \\ \overline{Y} \end{bmatrix} = \frac{1}{\overline{T^2} - \overline{T}^2} \begin{bmatrix} 1 & -\overline{T} \\ -\overline{T} & \overline{T^2} \end{bmatrix} \begin{bmatrix} \overline{TY} \\ \overline{Y} \end{bmatrix}$

$= \frac{1}{\overline{T^2} - \overline{T}^2} \begin{bmatrix} \overline{TY} - (\overline{T})(\overline{Y}) \\ (\overline{T^2})(\overline{Y}) - (\overline{TY})(\overline{T}) \end{bmatrix}$

12.11 Let $\mathbf{A} = \begin{bmatrix} 1 \\ m \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ giving us the least squares minimization problem $\arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2$. Let $\mathbf{x}^* = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{b} = \frac{x_0 + my_0}{1+m^2}$, so the point closest to $\mathbf{b}$ is $\begin{bmatrix} \frac{x_0+my_0}{1+m^2} \\ \frac{mx_0+m^2y_0}{1+m^2} \end{bmatrix}$

12.12 (a) Let $\mathbf{A} := \begin{bmatrix} \mathbf{x}_1^\top & 1 \\ \mathbf{x}_2^\top & 1 \\ \vdots & \vdots \\ \mathbf{x}_p^\top & 1 \end{bmatrix} \in \mathbb{R}^{p\times(n+1)}$, $\mathbf{z} := \begin{bmatrix} \mathbf{a} \\ c \end{bmatrix} \in \mathbb{R}^{n+1}$, and $\mathbf{b} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} \in$

$\mathbb{R}^p$.

This gives us the minimization problem $\arg\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{b}\|^2$.

(b) Let $\mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p] \in \mathbb{R}^{p\times n}$ and $\mathbf{c} = [1, 1, \ldots, 1]^\top \in \mathbb{R}^{p\times 1}$. It follows $\mathbf{A} = [\mathbf{X}\,\mathbf{c}]$. Thus, $\mathbf{z}^* = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$. Given $\sum_{i=1}^{p} \mathbf{x}_i = 0$,

$(\mathbf{A}^\top\mathbf{A})^{-1} = \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{c} \\ \mathbf{c}^\top\mathbf{X} & \mathbf{c}^\top\mathbf{c} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & 0 \\ 0 & p \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{X}^\top\mathbf{X})^{-1} & 0 \\ 0 & \frac{1}{p} \end{bmatrix} \in \mathbb{R}^{(p+1)\times(p+1)}$.

Given $\sum_{i=1}^{p} \mathbf{y}_i\mathbf{x}_i = 0$,

$\mathbf{A}^\top\mathbf{b} = \begin{bmatrix} \mathbf{X}^\top\mathbf{b} \\ \mathbf{c}^\top\mathbf{b} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{c}^\top\mathbf{b} \end{bmatrix} \in \mathbb{R}^{p+1}$.

Thus, $\mathbf{z}^* = \begin{bmatrix} (\mathbf{X}^\top\mathbf{X})^{-1} & 0 \\ 0 & \frac{1}{p} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{c}^\top\mathbf{b} \end{bmatrix} \in \mathbb{R}^{p+1} = \begin{bmatrix} 0 \\ \frac{1}{p}\mathbf{c}^\top\mathbf{b} \end{bmatrix} \in \mathbb{R}^{p+1}$.

Thus, the constant function $f(\mathbf{x}) = \frac{1}{p}\sum_{i=1}^{p} y_i$ is the affine function of best fit.

12.13 (a) Let $\mathbf{A} = [u_1, u_2 \ldots, u_n]^\top$ and $\mathbf{b} = [y_1, y_2, \ldots, y_n]^\top$.
Thus, $\hat{\theta_n} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} = \frac{\sum_{k=1}^{n} u_k y_k}{\sum_{k=1}^{n} u_k^2}$ using the least squares approach.

(b) If $u_k = 1$ for all $k > 0$, $\hat{\theta_n} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} u_k y_k}{\sum_{k=1}^{n} u_k^2} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n}(\theta + e_k)}{\sum_{k=1}^{n} 1}$

$= \lim_{n\to\infty} \frac{\theta n + \sum_{k=1}^{n} e_k}{n}$

$= \theta + \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} e_k = \theta \Leftrightarrow \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} e_k = 0$

12.14 Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}$.

Thus, $\mathbf{x}^* = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 18 \\ 11 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 3 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 18 \\ 11 \end{bmatrix} = \begin{bmatrix} \frac{7}{2} \\ \frac{1}{6} \end{bmatrix}$
gives us the least squares estimate.

12.22 $\|\mathbf{Ax} - \mathbf{b}_i\|^2 = \mathbf{x}^\top\mathbf{A}^\top\mathbf{Ax} - 2\mathbf{x}^\top\mathbf{Ab}_i + \|\mathbf{b}_i\|^2$.

Thus, $\sum_{i=1}^{p} \alpha_i \|\mathbf{Ax} - \mathbf{b}_i\|^2$

$= (\alpha_1 + \alpha_2 + \cdots + \alpha_p)\mathbf{x}^\top\mathbf{A}^\top\mathbf{Ax} - 2\sum_{i=1}^{p} \alpha_i\mathbf{x}^\top\mathbf{Ab}_i + \sum_{i=1}^{p} \alpha_i\|\mathbf{b}_i\|^2$

which is minimized when

$$
\begin{aligned}
\mathbf{x}^* &= \left((\alpha_1+\alpha_2+\cdots+\alpha_p)\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top(\alpha_1\mathbf{b}_1+\alpha_2\mathbf{b}_2+\cdots+\alpha_p\mathbf{b}_p)\\
&= \frac{1}{\alpha_1+\alpha_2+\cdots+\alpha_p}\sum_{i=1}^{p}\alpha_i\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\mathbf{b}_i\\
&= \frac{1}{\alpha_1+\alpha_2+\cdots+\alpha_p}\sum_{i=1}^{p}\alpha_i\mathbf{x}_i^*
\end{aligned}
$$

12.24 (a)

$$
\begin{aligned}
\mathbf{x}^{(0)} &= \left(\mathbf{A}_0^\top\mathbf{A}_0\right)^{-1}\mathbf{A}_0^\top\mathbf{b}^{(0)} = \mathbf{G}_0^{-1}\mathbf{A}_0^\top\mathbf{b}^{(0)}\\
\mathbf{x}^{(1)} &= \left(\mathbf{A}_1^\top\mathbf{A}_1\right)^{-1}\mathbf{A}_1^\top\mathbf{b}^{(1)} = \mathbf{G}_1^{-1}\mathbf{A}_1^\top\mathbf{b}^{(1)}
\end{aligned}
$$

(b) $\mathbf{G}_0 = \begin{bmatrix}\mathbf{A}_1^\top & \mathbf{a}_1\end{bmatrix}\begin{bmatrix}\mathbf{A}_1\\ \mathbf{a}_1^\top\end{bmatrix} = \mathbf{A}_1^\top\mathbf{A}_1 + \mathbf{a}_1\mathbf{a}_1^\top = \mathbf{G}_1 + \mathbf{a}_1\mathbf{a}_1^\top$

Thus, $\mathbf{G}_1 = \mathbf{G}_0 - \mathbf{a}_1\mathbf{a}_1^\top$

(c) $\mathbf{P}_1 = (\mathbf{G}_1)^{-1}$
$= \left(\mathbf{G}_0 - \mathbf{a}_1\mathbf{a}_1^\top\right)^{-1}$
$= \mathbf{G}_0^{-1} + \frac{(\mathbf{G}_0^{-1}\mathbf{a}_1)(\mathbf{a}_1^\top\mathbf{G}_0^{-1})}{1-\mathbf{a}_1^\top\mathbf{G}_0^{-1}\mathbf{a}_1}$
$= \mathbf{P}_0 + \frac{\mathbf{P}_0\mathbf{a}_1\mathbf{a}_1^\top\mathbf{P}_0}{1-\mathbf{a}_1^\top\mathbf{P}_0\mathbf{a}_1}$

(d) $\mathbf{A}_0^\top\mathbf{b}^{(0)} = \mathbf{G}_0\mathbf{G}_0^{-1}\mathbf{A}_0^\top\mathbf{b}^{(0)}$
$= \mathbf{G}_0\mathbf{x}^{(0)}$
$= (\mathbf{G}_1 + \mathbf{a}_1\mathbf{a}_1^\top)\mathbf{x}^{(0)}$
$= \mathbf{G}_1\mathbf{x}^{(0)} + \mathbf{a}_1\mathbf{a}_1^\top\mathbf{x}^{(0)}$

(e)

$$
\begin{aligned}
&\mathbf{x}^{(1)} = \mathbf{G}_1^{-1}\mathbf{A}_1^\top\mathbf{b}^{(1)}\\
&= \mathbf{G}_1^{-1}\mathbf{A}_1^\top\mathbf{b}^{(1)}\\
&= \mathbf{G}_1^{-1}(\mathbf{A}_1^\top\mathbf{b}^{(1)} + \mathbf{a}_1 b_1 - \mathbf{a}_1 b_1)\\
&= \mathbf{G}_1^{-1}(\mathbf{A}_0^\top\mathbf{b}^{(0)} - \mathbf{a}_1 b_1)\\
&= \mathbf{G}_1^{-1}(\mathbf{G}_1\mathbf{x}^{(0)} + \mathbf{a}_1\mathbf{a}_1^\top\mathbf{x}^{(0)} - \mathbf{a}_1 b_1)\\
&= \mathbf{x}^{(0)} + \mathbf{G}_1^{-1}\mathbf{a}_1\mathbf{a}_1^\top\mathbf{x}^{(0)} - \mathbf{G}_1^{-1}\mathbf{a}_1 b_1\\
&= \mathbf{x}^{(0)} - \mathbf{P}_1\mathbf{a}_1(b_1 - \mathbf{a}_1^\top\mathbf{x}^{(0)})
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbf{P}_{k+1} &= \mathbf{P}_k + \frac{\mathbf{P}_k\mathbf{a}_{k+1}\mathbf{a}_{k+1}^\top\mathbf{P}_k}{1-\mathbf{a}_{k+1}^\top\mathbf{P}_k\mathbf{a}_{k+1}}\\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - \mathbf{P}_{k+1}\mathbf{a}_{k+1}(b_{k+1} - \mathbf{a}_{k+1}^\top\mathbf{x}^{(k)})
\end{aligned}
$$

gives our update formula